Université de Genève

Département d'informatique

Théorie de l'information pour la science des données et l'apprentissage automatique

Série d'exercices 7 — 10.05.2023 Délai : 16.05.2023

Note : Veuillez suivre les instructions suivantes
`https://moodle.unige.ch/mod/page/view.php?id=171850`.

Attention, la base de tous les logarithmes est 2 !

$$H(X,Y) \triangleq \mathbb{E}_{p_{X,Y}}\left[-\log_2 p_{X,Y}(X,Y)\right] = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 p_{X,Y}(x,y),$$

$$I(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)},$$

$$D_{KL}(p||q) = \sum_x p(x)\log_2 \frac{p(x)}{q(x},$$

Problème 1 (20 points).
Prouvez que :

$$H(X,Y|Z) \geq H(X|Z)$$

Problème 2 (15 points).
Soient X et Y deux variables aléatoires avec une probabilité conjointe donnée dans le tableau suivant :

X	Y	$p_{X,Y}(x,y)$
0	0	1/8
0	1	1/4
1	0	1/8
1	1	1/4
2	0	0
2	1	1/4

- Calculez leur information mutuelle selon sa définition :

$$I(X;Y) = E_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$$

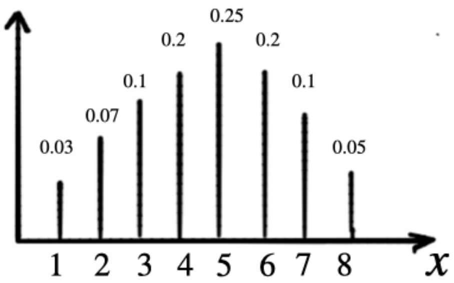
- Comparez votre résultat avec :

  (a) $I(X;Y) = H(X) - H(X|Y)$
  (b) $I(X;Y) = H(Y) - H(Y|X)$

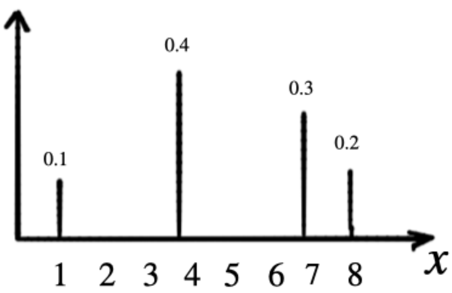
  et concluez sur l'équivalence des résultats obtenus.

Problème 3 (20 points).

- Ecrivez un programme Matlab/Python pour calculer la divergence KL entre deux distributions $p(x)$ et $q(x)$.
- Soit $p(x)$ une distribution **uniforme** définie sur l'intervalle $[1, 8]$. En utilisant votre code, calculez la divergence KL entre $p(x)$ et les cas suivants pour $q(x)$:



(a)



(b)

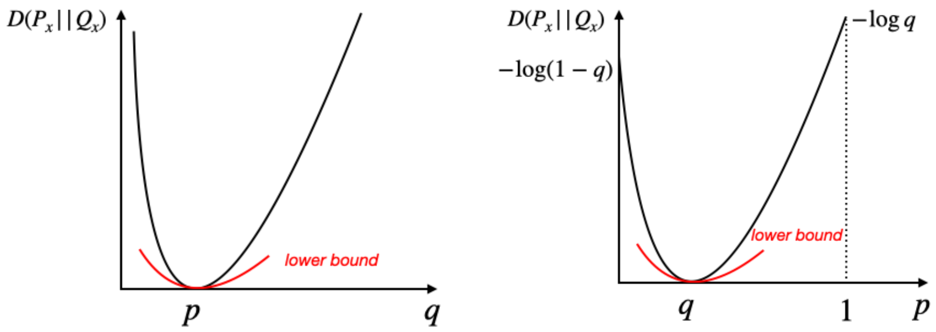
Remarque : Attention aux divisions par zéro.

Problème 4 (Divergence binaire — 25 points).

Soit la variable aléatoire binaire $X \in \{0,1\}$. La divergence binaire entre deux distributions $P_x = \{p, 1-p\}$ et $Q_x = \{q, 1-q\}$ est définie par :

$$D(P_x||Q_x) = d(p,q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$$

La quantité $d(p,q)$ dépend de p et q ainsi :



(a) (b)

Figure 2: Divergence binaire

1. Prouvez les propriétés suivantes analytiquement :
   i. $d(p,q) = 0 \quad$ pour $\quad q = p$
   ii. $d(p,q) = -\log(1-q) \quad$ pour $\quad p = 0$
   iii. $d(p,q) = -\log(q) \quad$ pour $\quad p = 1$
2. Ecrivez un code Matlab/Python pour visualiser la Fig. 2 pour
   a) $p = 0.1$ et $q \in [0,1]$
   b) $q = 0.2$ et $p \in [0,1]$
3. Il existe une borne inférieure donnée par la forme quadratique suivante :

$$d(p,q) \geq 2(p-q)^2 \log e$$

   Visualisez cette borne inférieure sur vos graphiques précédents, d'une couleur différente.

Problème 5 (20 points).

Prouvez l'égalité :

$$\arg\min_{q(x)} H(p(x), q(x)) = \arg\min_{q(x)} D_{KL}(p(x)||q(x))$$

Remarque : Cette égalité montre l'équivalence entre la maximisation de la vraisemblance et la minimisation de la divergence KL par rapport à $q(x)$.

Note : Ici nous travaillons avec l'entropie croisée $H(p,q) = -\sum_x p(x) \log q(x)$